

Data-Driven Sensitivity Inference for Thomson Scattering Electron Density Measurement Systems

Keisuke Fujii,¹ Ichihiko Yamada,² and Masahiro Hasuo¹

¹*Department of Mechanical Engineering and Science,
Graduate School of Engineering, Kyoto University, Kyoto 615-8540, Japan*

²*National Institute for Fusion Science, Toki 509-5292, Japan*

(Dated: December 13, 2016)

We developed a method to infer the calibration parameters of multichannel measurement systems, such as channel variations of sensitivity and noise amplitude, from experimental data. We regard such uncertainties of the calibration parameters as dependent noise. The statistical properties of the dependent noise and that of the latent functions were modeled and implemented in the Gaussian process kernel. Based on their statistical difference, both parameters were inferred from the data.

We applied this method to the electron density measurement system by Thomson scattering for Large Helical Device plasma, which is equipped with 141 spatial channels. Based on the 210 sets of experimental data, we evaluated the correction factor of the sensitivity and noise amplitude for each channel. The correction factor varies by $\approx 10\%$, and the random noise amplitude is $\approx 2\%$, i.e., the measurement accuracy increases by a factor of 5 after this sensitivity correction. The certainty improvement in the spatial derivative inference was demonstrated.

I. INTRODUCTION

Multichannel measurement systems have become more popular in scientific communities because of the reduced cost of digitizers, sensors, and signal processors. In a multichannel system, sensitivity and noise amplitude variations over channels are important calibration parameters. Usually, these parameters are evaluated by separate calibration experiments. However, achieving a highly accurate calibration is sometimes difficult because of the calibration principle or temporal variation of these parameters. If the sensitivity uncertainty exceeds the signal uncertainty by random statistics, then the signals will suffer from scatters that are not independent in multiple measurements.

The data observed by a measurement system contains information about the system itself [1]. The upper panel in Fig. 1 (a) shows two examples of the data observed by one of multichannel systems, specifically, the electron density (n_e) spatial distribution measurement in the high-temperature plasma (the details of which are provided below). Some data points have larger or smaller values than those in the vicinity for the two measurements. This dependent scatter is likely due to the miscalibration of the system sensitivity, and it significantly reduces measurement accuracy. Humans can clearly distinguish this dependent noise quantitatively by finding the correlation in multiple observation results. In this study, we demonstrated the machine learning of such dependent noise using the Gaussian process regression (GPR) framework [2, 3].

GPR is a statistical model that can handle the probability distribution on functions. The main advantages of GPR are as follows:

1. GPR is non-parametric, i.e., a specific function form is not assumed.
2. GPR is fully based on Bayesian statistics, i.e., if the model complexity is appropriately parameterized, then over- and under-fitting can be avoided.

Given the above advantages, GPR is being adopted in a wide range of fields that involve scientific data analysis, such as life sciences [4, 5], chemistry [6, 7], and astrophysics [8, 9]. GP has also been used to infer the latent function and its derivatives from data generated by the Thomson scattering (TS) system, which will also be analyzed in this work [10, 11]. However, simple noise has been often assumed for the analysis, e.g., independent homoscedastic Gaussian noise or Poisson noise. Few works have been reported regarding dependent noise inference [1], which is important for the real-world measurement systems.

In this work, we modeled the statistical properties of dependent noise such as sensitivity variation or channel-dependent noise amplitude into the kernel of Gaussian process (GP). Using these statistical models, we show that the inference of these noises is straightforward in the GPR framework. We also applied this technique to the TS system for n_e spatial distribution measurement in the Large Helical Device (LHD-TS system) [12, 13], an example of which is shown in Fig. 1(a).

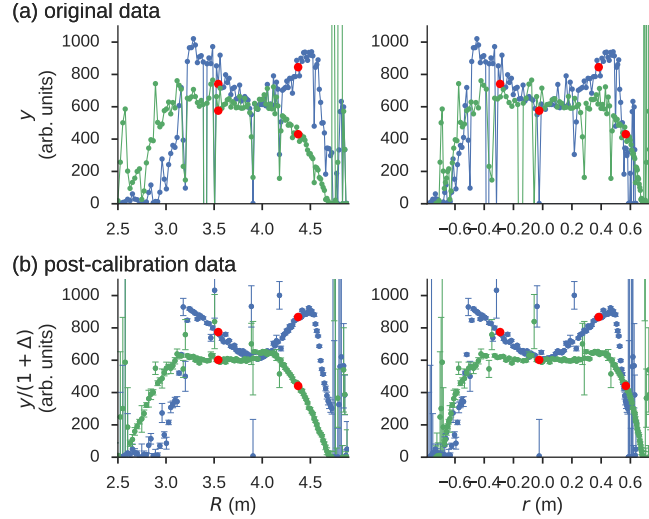


FIG. 1. (a) Two sets of the data provided by the LHD-TS system. Green points show the data observed for shotnumber #120725 at $t = 4.2$ s and blue ones show that for #120763 at $t = 5.6$ s. Some measurement points have a smaller or a larger value than the vicinity points for both results. The left panel is in the original coordinate (major radius R), whereas the right panel is in the mapped coordinate (minor radius r). To clarify the detector channel positions, we show the data points measured by the detector $i = 50$ and 100 in red markers. (b) Calibrated results with a misalignment factor Δ are inferred in this work. Most scatters were removed after the calibration.

The rest of this paper is organized as follows. In the next section, we briefly describe the fundamental usage of GPR for the latent function and noise inference. The essence of the inference is based on the statistical difference between the latent function and noise. Therefore, their statistical modeling is important. We slightly extended the theory in order to treat the dependent noise for the multichannel measurement systems. In section III, we apply this method to the experimental data obtained by the LHD-TS system. An overview of the system and the statistical model of the noise are described in this section. The application results are shown in subsection IIIB, and the effect of this method is presented in section IIIC, in particular, in the spatial derivative inference.

II. DEPENDENT NOISE ESTIMATION BASED ON GP

A. Fundamentals of GP

GP is probability distribution on function space. If function f follows GP with mean 0 and kernel K_f , then the joint distribution of any realizations $\mathbf{f} = \{f_i = f(x_i) | i \in 1, \dots, N\}$ of function f is

$$p(\mathbf{f}) = \mathcal{N}(0, K_f), \quad (1)$$

where $\mathcal{N}(\mu, \Sigma)$ is the multivariate normal distribution with mean μ and covariance Σ [2, 3, 14].

Additive and independent noise \mathbf{n} is frequently assumed, where the observation data $\mathbf{y} = \{y_i | i \in 1, \dots, N\}$ are the sum of the function values and noise, $\mathbf{y} = \mathbf{f} + \mathbf{n}$. If both \mathbf{f} and \mathbf{n} follow zero mean GP with respective kernels K_f and K_n , then the observation and latent function values jointly follow zero mean GP with

$$p\left(\begin{bmatrix} \mathbf{y} \\ \mathbf{f} \end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} K_f + K_n & K_f \\ K_f & K_f \end{bmatrix}\right). \quad (2)$$

For the latent function kernel K_f , similarity in the vicinity is often assumed (the meanings of *similar* and *vicinity* may depend on problems). One of the most popular kernels is the radial basis function (RBF) kernel [2, 14],

$$K_f(i, i') = \alpha \exp\left[-\frac{(x_i - x_{i'})^2}{2l^2}\right] \quad (3)$$

where $\mathbf{x} = \{x_i | i \in 1, \dots, N\}$ is the explanatory variable (usually, spatial coordinates for the multichannel measurement) and l is the scale length. With this kernel, function values at two close locations, $|x_i - x_{i'}| \ll l$, have a significant correlation of approximately 1, whereas those at distant locations, $|x_i - x_{i'}| \gg l$, have no correlation. $\alpha > 0$ exhibits self-variation, which indicates the variation of the function values around zero.

The GP kernel of the additive independent Gaussian noise is written as

$$K_n(i, i') = \sigma^2 \delta(i - i') \quad (4)$$

where σ is the noise variance, and $\delta(i)$ is the Dirac delta function.

A simple algebraic calculation of Eq. 2 obtains the posterior density of the function values \mathbf{f} and noise \mathbf{n} with given data \mathbf{y} as

$$p(\mathbf{f}|\mathbf{y}) = \mathcal{N}\left(K_f(K_f + K_n)^{-1}\mathbf{y}, K_f - K_f(K_f + K_n)^{-1}K_f\right), \quad (5)$$

and

$$p(\mathbf{n}|\mathbf{y}) = \mathcal{N}(\mathbf{K}_n(\mathbf{K}_f + \mathbf{K}_n)^{-1}\mathbf{y}, \mathbf{K}_n - \mathbf{K}_n(\mathbf{K}_f + \mathbf{K}_n)^{-1}\mathbf{K}_n), \quad (6)$$

respectively. This inference is based on the statistical difference of the latent function and noise, which are implemented in the GP kernels, as in Eqs. 3 and 4. Note that the notation $p(\mathbf{f}|\mathbf{y})$ represents the conditional probability distribution of the variable \mathbf{f} conditioned by \mathbf{y} .

The set of hyperparameters $\boldsymbol{\theta}$ (α , l , and σ in this case) can be estimated by the type-II maximum posterior (MAP) method, or they are numerically marginalized out. In the type-II MAP method, the point estimate of $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\theta} \approx \operatorname{argmax}_{\boldsymbol{\theta}} [\log p(\boldsymbol{\theta}|\mathbf{y})] \quad (7)$$

where $p(\boldsymbol{\theta}|\mathbf{y})$ is the marginal posterior density for $\boldsymbol{\theta}$ with given data \mathbf{y} , where \mathbf{f} or \mathbf{n} is marginalized out

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \quad (8)$$

where $p(\boldsymbol{\theta})$ is the prior distribution of the hyperparameters (hyperprior). The log of the marginalized likelihood $p(\mathbf{y}|\boldsymbol{\theta})$ can be written as follows [2, 14]:

$$\log p(\mathbf{y}|\boldsymbol{\theta}) = \log \int p(\mathbf{y}, \mathbf{f}|\boldsymbol{\theta}) d\mathbf{f} = -\frac{1}{2} \log |\mathbf{K}_f + \mathbf{K}_n| - \frac{1}{2} \mathbf{y}^\top (\mathbf{K}_f + \mathbf{K}_n)^{-1} \mathbf{y} - \frac{N}{2} \log(2\pi). \quad (9)$$

B. Common-Mode Noise Inference

In the previous subsection, we separated the observation data into the latent function component and noise component, as shown in Eqs. 5 and 6, on the basis of their statistical difference. Such separation is not restricted to the latent function and noise pair. If data are the sum of multiple components that have different statistical properties and they are appropriately implemented into the GP kernels, then the posteriors of these elements can be obtained by an equation similar to Eq. 5 or 6.

An example is the common-mode additive noise \mathbf{n}_C , which takes the same values among multiple sets of data (indexed by $j \in 1, \dots, M$). Let us consider that data \mathbf{y} are the sum of the latent function values \mathbf{f} , the random noise that is independent in each data set \mathbf{n}_R , and the random common-mode noise \mathbf{n}_C .

The kernel of \mathbf{n}_R for M sets of data can be written as

$$\mathbf{K}_{n_R}(i, j, i', j') = \sigma_R^2 \delta(i - i') \delta(j - j') \quad (10)$$

and that for \mathbf{n}_C is

$$K_{\mathbf{n}_C}(i, j, i', j') = \sigma_C^2 \delta(i - i') \quad (11)$$

If the latent functions in the j -th data set \mathbf{f}_j are independent from that in another data set, then the kernel of the joint distributions is block diagonal

$$K_f = \begin{bmatrix} K_{f_1} & 0 & \dots & 0 \\ 0 & K_{f_2} & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & K_{f_M} \end{bmatrix}. \quad (12)$$

where K_{f_j} are the kernels for the latent function in the j -th data set.

We assume that all the noise can be approximated to be additive and independent of each other. Thus, the kernel of the observations is written as $K_y = K_f + K_{\mathbf{n}_R} + K_{\mathbf{n}_C}$. By repeating a discussion similar to Eqs. 5 and 6, we can analytically evaluate the posterior densities of \mathbf{f} , \mathbf{n}_R , and \mathbf{n}_C .

Note that the independent assumption of \mathbf{f}_j is too strong because some similarity usually exists among latent functions in multiple data sets. To protect an artifact from this assumption, we analyze a few ($M = 6$ in this work) data sets simultaneously. To analyze more data, the nondiagonal kernel that represents the similarity needs to be introduced. Although a bit more details will be discussed in the final section, it is left for the future study and we focus on the principle demonstration in this work.

C. Noise Modeling for MultiChannel Systems

In this subsection, we model other realistic dependent noises that are important for multichannel measurement systems.

1. Miscalibration Noise

The sensitivity uncertainty in the multichannel measurement system can be viewed as multiplicative common-mode noise. Here we define the miscalibration factor for channel i as Δ_i , where the true calibration factor R_i and the calibration factor that we obtained in the calibration experiment R_i^0 are presented in the following relation:

$$R_i = R_i^0(1 + \Delta_i). \quad (13)$$

We assume Δ_i is independent of each other and distributed around zero. One of the prior candidates for Δ_i is the homoscedastic normal distribution

$$\Delta_i \sim \mathcal{N}(0, \sigma_\Delta^2). \quad (14)$$

The noise due to the miscalibration factor is written as $n_{\Delta i,j} = \Delta_i f_{i,j}$. Therefore, the kernel of this noise K_Δ can be approximated as

$$K_{n_\Delta}(i, i', j, j') = \sigma_\Delta^2 \hat{f}_{i,j} \hat{f}_{i',j'} \delta(i - i'), \quad (15)$$

where $\hat{f}_{i,j}$ is an estimate of the latent function value. We note that Eq. 15 is an approximation because this noise is not strictly additive. If the posterior width of $f_{i,j}$ is larger than compared to its mean value, then this approximation becomes invalid.

2. Channel Dependent Noise

In real-world multichannel systems, noisy channels that produce more scattered signals than other channels may exist. We modeled a channel-dependent noise, \mathbf{n}_D , such that the noise variations σ_{Di} depends on channel i , while the noise itself varies with zero mean Gaussian distribution. The kernel of this noise can be written as

$$K_{n_D}(i, i', j, j') = \sigma_{Di}^2 \delta(i - i') \delta(j - j'), \quad (16)$$

Application of a prior distribution for the noise variances, $p(\sigma_{Di}^2)$, may be reasonable .

III. APPLICATION TO LHD-TS SYSTEM DATA

A schematic illustration of the LHD-TS system is shown in Fig. 2 [13]. The Nd:YAG laser is injected into the plasma, and the scattered light is focused on the edges of an optical fiber array that comprises 141 fibers [Fig. 2 (a)]. The scattering light comes only from the cross-point between the laser path and the sight line, and therefore this diagnostics system provides high spatial resolution (typically, on the order of a few centimeters). The temporal resolution is determined by a laser pulse width of (≈ 10 ns), which is much shorter than the time scale of the plasma turbulence. Given its high spatial and temporal resolution, the TS system has been installed in almost all major magnetic plasma confinement devices [12, 15–18].

The light is transferred for ≈ 50 m to a set of polychrometers [Fig. 2 (b)] in the diagnostics room. Each polychrometer is equipped with six interference filters and avalanche photodiodes (APD). The

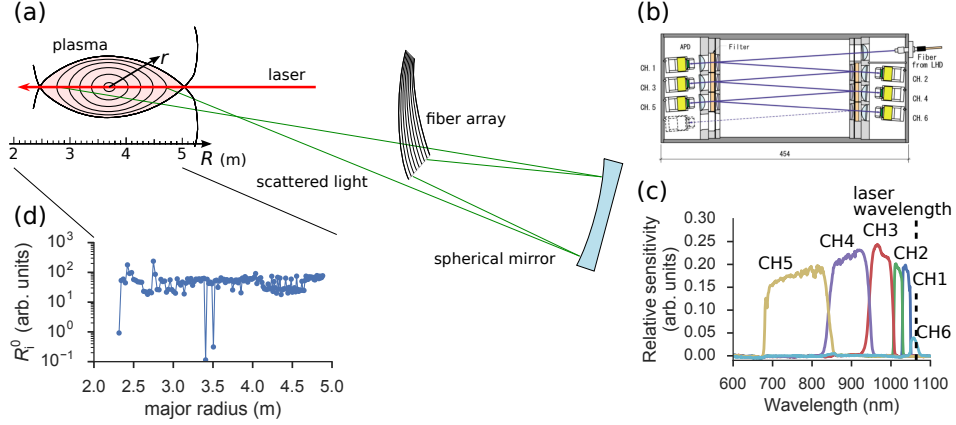


FIG. 2. Schematic illustration of the LHD-TS system. (a) Optical configuration of LHD plasma, the incident laser, and an optical system for collecting the scattering light. The nested magnetic surface, the major radius (R), and minor radius (r) in the plasma are also shown. (b) A polychrometer for observing the scattering light spectrum (quoted from Ref. [13]). (c) The relative wavelength sensitivities for the polychrometer measured by the first-step calibration experiment. (d) The absolute throughput for the LHD-TS system measured by the second-step calibration experiment.

spectral profile of the scattered light is estimated by five channels (CH1–CH5). The sixth APD (CH6) measures the light around the laser wavelength and is not used for the measurement because the stray laser light inside the vacuum chamber significantly contaminates this wavelength.

One spectrum is observed by one polychrometer. Therefore, 141 polychrometers are used for the spatial profile measurement. Although plasma emission is also detected by these detectors, the intensity of plasma emission is estimated by the subsequent measurement just after the laser pulse and subtracted from the signal [13].

Two-step calibration was conducted for the LHD-TS system. For the first step, the wavelength-dependent sensitivity of each polychrometer (including the transmittance of the interference filters and sensitivities of the APDs) is calibrated against a standard halogen lamp [13]. An example of the results of the wavelength-dependent sensitivity is shown in Fig. 2 (c). For the second step, the absolute throughput from the scattering position to the APDs is calibrated by Raman and Rayleigh scattering experiments, which are performed once for each experimental campaign [12, 19]. The results of the absolute calibration factor R_i^0 obtained in the 2013 experimental campaign is shown in Fig. 2 (d). The variation of the calibration factor is almost of an order of magnitude.

The measurement points at each major radius R are mapped on the basis of T_e results into the nested magnetic flux coordinate, which is denoted by minor radius r [see Fig.2 (a)] [20]. The mapped results are shown in the right panels in Fig. 1. The parameter profiles, including n_e , are

widely accepted to become axisymmetric in the mapped coordinate.

The random independent noise in the LHD-TS data arises because of the finite number of detected photons (shot noise) and thermal noise of the APDs and digitizers, as well as the fluctuation of the plasma emission. The dependent noise in the LHD-TS data is due to the miscalibration of the system and the existence of faulty detectors such as those that have bad contact in electric circuits.

A. Statistical Modeling of the LHD-TS System

In this subsection, we describe our modeling of the latent function of n_e and the noise sources for the LHD-TS system. For the latent function kernel, K_f , we adopted the RBF kernel (Eq. 3) and modified it in order to consider the prior information that the n_e profile in the magnetic plasma confinement device is axisymmetric, as follows:

$$K_f(i, i', j, j') = \alpha_j \left(\exp \left[-\frac{(r_{i,j} - r_{i',j'})^2}{2l_j^2} \right] + \exp \left[-\frac{(r_{i,j} + r_{i',j'})^2}{2l_j^2} \right] \right) \delta(j - j'). \quad (17)$$

where $\alpha_j > 0$ and $l_j > 0$ are hyperparameters for this kernel. The first term of Eq. 17 indicates a smooth n_e profile, i.e., the correlation between two n_e values that are closer to each other is larger. The second term makes two n_e values at the other side against the magnetic axis ($r = 0$) close. The flat noninformative priors are used for α_j and l_j .

We consider the following three types of quasi-additive noise for the LHD-TS data:

1. Common-mode multiplicative noise, \mathbf{n}_Δ .
2. Independent Gaussian noise with a variation that is dependent on channel \mathbf{n}_D .
3. Independent Gaussian noise with a variation proportional to the latent function value, \mathbf{n}_P .

We modeled the kernel of \mathbf{n}_Δ similar to Eq. 15, but with a hierarchical form,

$$K_{\mathbf{n}_\Delta}(i, i', j, j') = \sigma_{\Delta i}^2 \hat{f}_{i,j} \hat{f}_{i',j'} \delta(i - i'), \quad (18)$$

with a hyperprior on $\sigma_{\Delta i}^2$,

$$\sigma_{\Delta i}^2 \sim \mathcal{IG}(\frac{1}{2}, \frac{\bar{\sigma}_\Delta^2}{2}). \quad (19)$$

where $\mathcal{IG}(\alpha, \beta)$ is the inverse gamma distribution with a shape parameter α and scale parameter β . Moreover, the variance of the hyperprior $\bar{\sigma}_\Delta^2$ is another hyperparameter. Once the variance $\sigma_{\Delta i}$ is marginalized out, the distribution of this noise becomes Cauchy distribution with scale $\bar{\sigma}_\Delta$.

For the GP kernel of \mathbf{n}_D , we adopted Eq. 16 also with the prior distribution on σ_{Di}

$$\sigma_{Di}^2 \sim \mathcal{IG}(\frac{1}{2}, \frac{\bar{\sigma}_D^2}{2}). \quad (20)$$

We modeled the third noise \mathbf{n}_P such that its variation is proportional to the latent function value. The GP kernel for this noise can be approximated by

$$K_{nP}(i, i', j, j') = \sigma_P^2 \hat{f}_{i,j}^2 \delta(i - i') \delta(j - j'). \quad (21)$$

The above modeling is considerably simplified, especially, for low T_e plasmas. The spectral profile of the scattered light from a low T_e plasma, which is usually located at the edges of plasma, is close to the laser spectrum. The CH6 APD in Fig. 2(c) cannot be used for the measurement due to the strong stray light. Thus, the evaluation of T_e and n_e values from the small T_e plasma becomes unstable, and the evaluated values sometimes deviate significantly from the true value.

In our modeling, this T_e dependence of the signal variance is not considered. We model only the channel dependence of the variation. Thus, the noise variation of the channels observed on the edges of plasma is inferred to be very large. However, we focus on the evaluation of the miscalibration factor in this work.

The point estimates of the hyperparameters in these kernels were conducted by the type-II MAP method (Eq. 8). Because kernels contain $\hat{f}_{i,j}$, we performed an iterative evaluation until the solutions converged.

B. Results

We prepared 210 sets of LHD-TS data (from two LHD discharges, #120724 and #120762) in the 2013 experimental campaign. The number of spatial channels N is 141. The radial coordinate r is mapped based on the basis of VMEC calculation [20].

We randomly divided 210 data set into 35 mini-batches where in each mini-batch we analyzed $M = 6$ data sets at once. The results were combined with a Bayesian committee machine [21],

$$p(\mathbf{n}_\Delta | \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(D)}) \propto \prod_{d=1}^D \frac{p(\mathbf{n}_\Delta | \mathbf{y}^{(d)})}{p(\mathbf{n}_\Delta)^{D-1}}, \quad (22)$$

where $\mathbf{y}^{(d)}$ is the d -th mini-batch indexed by $d = 1, \dots, D$. In this work, $D = 35$. Note that

Such mini-batch calculation is adopted not only to relax the large calculation cost of GP, i.e., $\mathcal{O}((N \times M)^3)$, but also to avoid the strong assumption in Eq. 12.

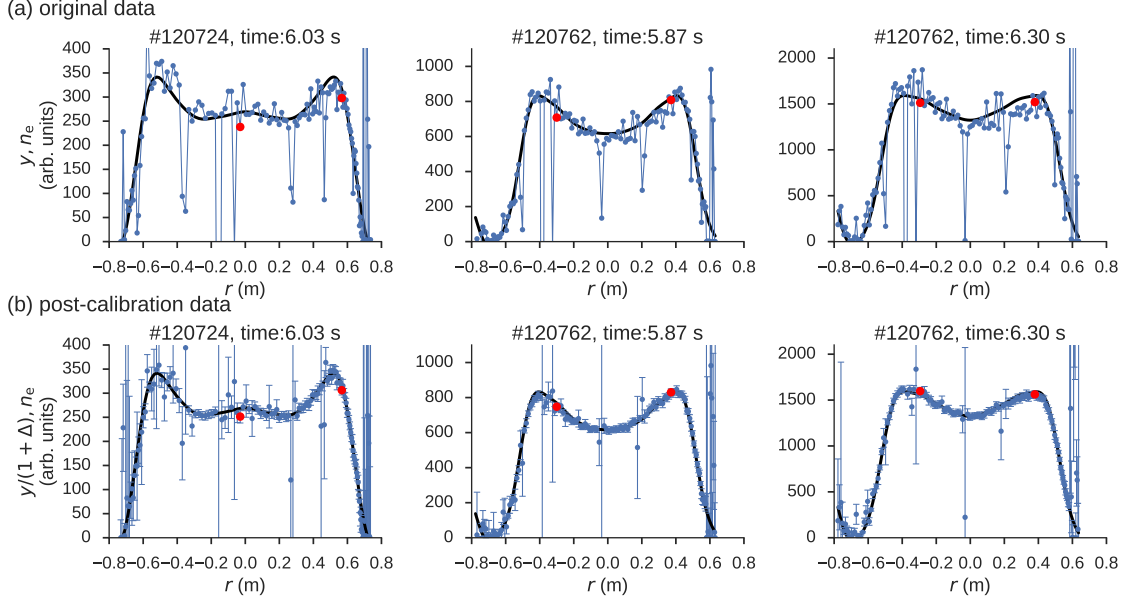


FIG. 3. (a) Examples of the data \mathbf{y} obtained by the LHD-TS system as a function of effective minor radius r . To show the mapping effect for each data set, the data obtained by channels $i = 50$ and 100 are shown in red. (b) Calibrated values, $y_i/(1 + \Delta_i)$. Scatters of the signals are significantly removed. (Black curves) Posterior mean of the latent functions $f_{i,j}$.

Figure 3 (a) shows some of the experimental data used for one mini-batch. We maximized the marginalized posterior conditioned by this data set (Eq. 9). Moreover, σ_{Δ_i} and σ_{D_i} that maximize the posterior are shown in Fig. 4. Higher σ_{D_i} values were evaluated for possible faulty channels, e.g. $i = 43$, and both ends of the plasma $i < 30$ and $i > 123$ where the T_e values are always low. $\bar{\sigma}_{\Delta}$, $\bar{\sigma}_D$, and σ_P were evaluated as 0.075, 9.3, and 0.018, respectively.

With these hyperparameters, we evaluated the posterior distributions for the latent functions for n_e and the miscalibration factor Δ_i based on Eqs. 5 and 6, respectively. The posteriors of n_e and Δ_i are shown in Fig. 3 by black curves and in Fig. 5 by blue points with shadows that represent their one standard deviations, respectively.

Based on a relation $y_{i,j} = f_{i,j} + n_{D_{i,j}} + n_{P_{i,j}} + \Delta_i \hat{f}_{i,j}$, the following calibration can be made with the given Δ_i values:

$$\frac{y_{i,j}}{1 + \Delta_i} = f_{i,j} + \frac{n_{D_{i,j}} + n_{P_{i,j}}}{1 + \Delta_i}. \quad (23)$$

The mean and standard deviation of $y_{i,j}/(1 + \Delta_i)$ are shown in Fig. 3 (b) by blue markers and error bars. The scatters of the data points are largely suppressed by this post-calibration.

The upper and lower panels of Fig. 6 show the posterior mean of Δ_i and σ_{D_i} for the 35 mini-batches, respectively. These results have similar values. The green line in the upper panel of Fig.

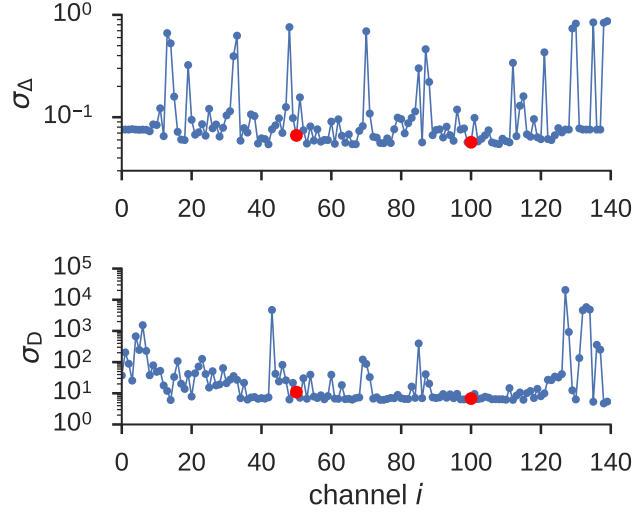


FIG. 4. The variance of miscalibration noise σ_{Δ_i} and the random noise variance σ_{D_i} that maximize the marginal posterior Eq. 7. Results for channels $i = 50$ and 100 are shown in red.

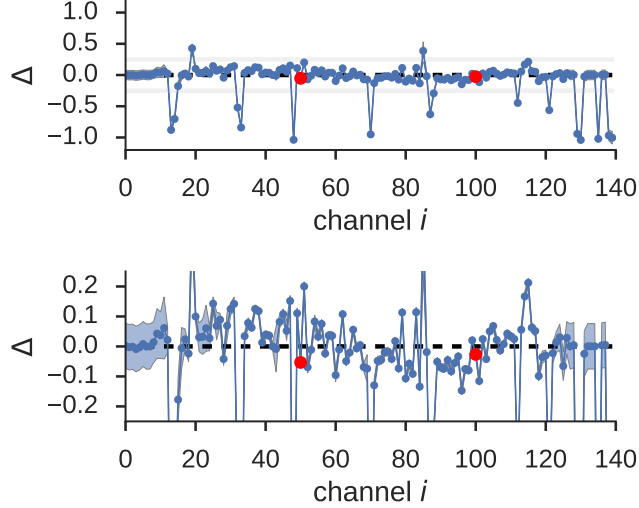


FIG. 5. Miscalibration noise Δ_i inferred from one mini-batch. The posterior mean is shown by markers and the range within one standard deviation are shown with shadow. The values for channels $i = 50$ and 100 are shown in red. The lower panel is the vertical expansion of the upper panel.

6 shows the combined results of Δ_i by the Bayesian committee machine, while that in the lower panel is the median of these results of σ_{D_i} . Note that the Bayesian committee machine is not used to combine multiple results of σ_{D_i} since we only make a point estimate of this hyperparameter rather than its posterior distribution.

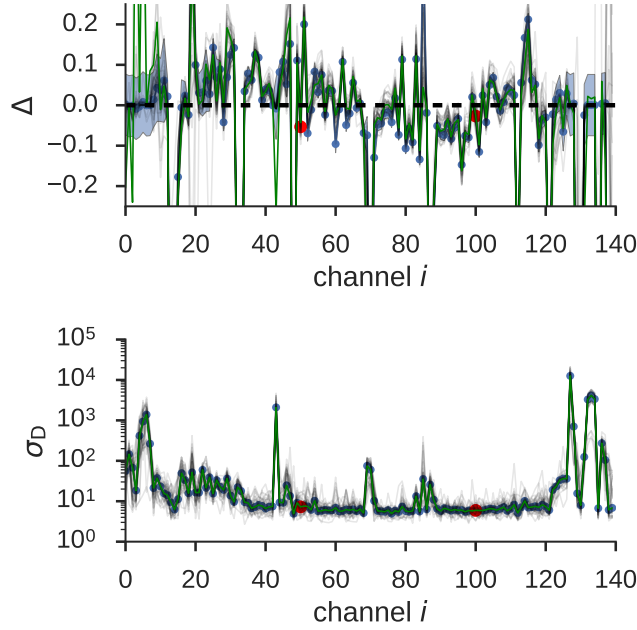


FIG. 6. (Upper panel) Light gray curves show the posterior means of Miscalibration noise Δ for all the data sets. The green curves show the combined result by the Bayesian committee machine. Blue markers and blue shadow are the evaluated results for one small data set, which is the same as those in Figs.5 (Lower panel) Light gray curves show point estimates of random noise variance σ_D , for all the data sets. The green curve is the median of them. Blue markers are the evaluated results for one small data set, which is the same as those in Figs. 4.

With these combined results, we calibrated test data that are not used for the above inference. The calibration was performed on the basis of Eq. 23. Some of the original data and the results calibrated with Δ_i values (post-calibration) are shown in Figs. 1 (a) and (b), respectively. Most scatters in the test data were also removed, thereby indicating that our method successfully avoids overfitting. A few outliers still exist in the post-calibrated data, and they are likely due to the simplified modeling for the outliers.

C. Spatial Derivative Inference

We demonstrate a benefit of this post-calibration by evaluating the latent function and its first and second spatial derivatives, which are important for the plasma simulations and transport analysis [22].

A original test data set analyzed here is shown in the left panel of Fig. 7 (a) which is also shown

by the green markers in Fig. 1 (a). The post-calibration was performed on the basis of Eq. 23 for this test data set, and the result is shown in the right panel.

To analyze the original data (without the post-calibration), we adopted a basic GP regression

$$K = K_f + K_{n_R}, \quad (24)$$

as described in subsection II A. The basic GP regression cannot deal with the outliers. We manually removed the outliers from the analysis as shown by green markers in Fig. 7 from the analysis.

For the post-calibration data, we adopted the following kernel,

$$K = K_f + K_{n_D} + K_{n_P}, \quad (25)$$

where the hyperparameters in K_{n_C} and K_{n_P} are fixed to those evaluated for the training data multiplied by $1/(1 + \Delta_i)$. We have already inferred outliers as signal variation σ_{D_i} . Manual removal of the outliers is not necessary for the post-calibration data.

Here we carried out Markov chain Monte-Carlo (MCMC) method to marginalize all the hyperparameters. The inferred latent functions are shown by black curves in Fig. 7 (a) for both cases. The scale length of the RBF kernel (l in Eq. 17) that maximizes the mearginal posterior is shown by the horizontal bar. The ranges within two standard deviation of the posterior are shown in gray shadow. The optimum scale length for the original test data in Fig. 7 is larger than that for the post-calibrated data, resulting in lower spatial resolution for the inference.

We also evaluated the spatial derivatives against the minor radius r by GP [2, 11]. The inferred first and second derivatives are shown in (b) and (c) in Fig. 7, respectively. The spatial resolution (inverse of the scale length) or the certainties of the derivative evaluation also increase by the post-calibration. The spatial resolution (inverse of the scale length) or the certainties of the derivative evaluation also increase by the post-calibration.

IV. SUMMARY AND FUTURE PROSPECTS

We developed a method to infer the miscalibration factor and channel dependent noise amplitude for multichannel measurement systems based on GPR. In this method, the dependent noises are statistically modeled in the kernel that is hierarchically parameterized. These hyperparameters were estimated so that they jointly maximize the marginal posterior. The posterior distributions of the miscalibration factor were inferred with these hyperparameters. We applied this method to the TS system for n_e measurement in LHD. We showed that the post-calibration with these mis-

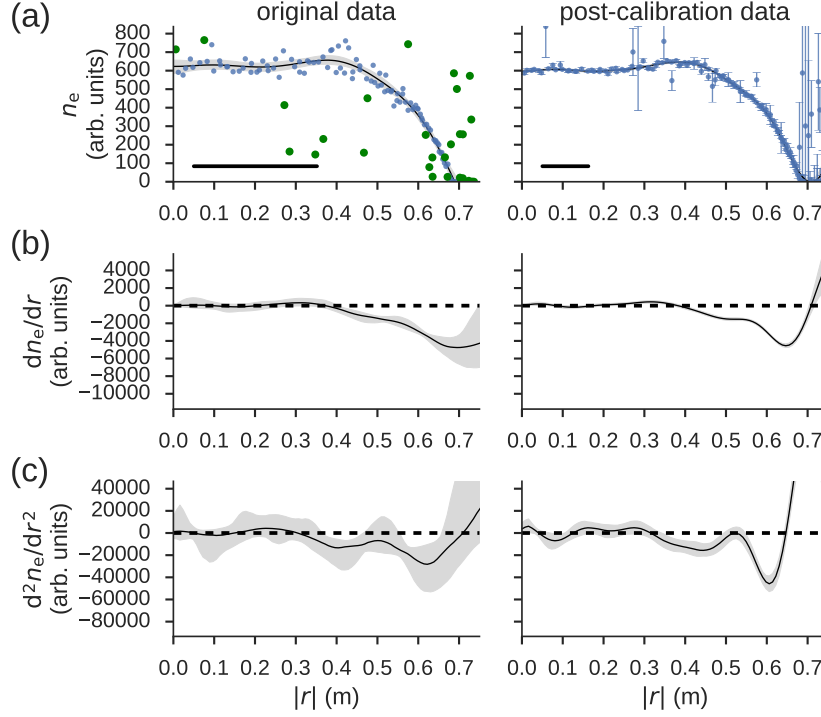


FIG. 7. Analyzed results for the test data (shot number = 120725, $t = 4.2$ s). (a) The left panel shows the original data measured by the LHD-TS system. The green markers are outliers that are chosen manually and removed for the inference. The right panel shows the post-calibrated data. The posterior median of the inferred latent functions by MCMC are shown by black curves. The horizontal bar in each panel shows the length scale in the RBF kernel that maximizes the marginal posterior. (b) First and (c) second spatial derivatives of the latent functions. The range within 95% intervals of the posterior is shown with shadows in each panel.

calibration factors and channel-dependent noise amplitude significantly decreases the uncertainty, especially, for spatial derivative estimation.

In the present stage of analysis, we adopted the mini-batch calculation with $M = 6$, to avoid the artifact due to the too simple assumption, Eq.12. Equation 12 is not valid for analyzing many sets of data, since this assumes that M sets of the data are random samples from $\mathcal{N}(0, K_f)$, while there are some clear similarities (or correlations) among them, e.g. all the n_e profiles are positive-valued and have larger values in the plasma axis than that at the edge region. In the future analysis, it is necessary to adopt more appropriate models that distinguishes the correlation among latent functions and the dependent noise. A possible model that takes the correlation of the latent function into account is the coregionalized model [23], which models non-diagonal covariance matrix.

Another future improvement is to consider the T_e dependence of the noise variance, which we neglected in this work. Due to the spectroscopic configuration of the LHD-TS system, the measurement accuracy for n_e decreases in very low or high T_e plasmas. The measurement accuracy for T_e also depends on both T_e and n_e values. Such correlation should be taken into account to infer the latent functions in wide range of plasma parameters.

ACKNOWLEDGMENTS

The authors thank the development team of pyGPs, the Python implementation of GP. The authors would also like to thank the LHD Experiment Group for the smooth LHD operation. Part of this work was supported by JSPS KAKENHI grant number 26610191, the National Institute for Fusion Science (NIFS14KLPH023, NIFS13KLPH032).

-
- [1] N. P. Gibson, S. Aigrain, S. Roberts, T. M. Evans, M. Osborne, and F. Pont, Monthly Notices of the Royal Astronomical Society **419**, 2683 (2012), <http://mnras.oxfordjournals.org/content/419/3/2683.full.pdf+html>.
 - [2] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, US One Rogers Street Cambridge MA 02142-1209, 2005).
 - [3] S. H. S. D. D. B. V. A. R. D. B. Gelman A., Carlin J. B., *Bayesian Data Analysis 3rd edn* (CRC Press, 6000 Broken Sound Pkwy NW 300, Boca Raton, FL 33487, 2013).
 - [4] O. Stegle*, S. V. Fallert, D. J. C. MacKay, and S. Brage, IEEE Transactions on Biomedical Engineering **55**, 2143 (2008).
 - [5] M. Kemmler, E. Rodner, P. Rsch, J. Popp, and J. Denzler, Analytica Chimica Acta **794**, 29 (2013).
 - [6] F. R. Burden, Journal of Chemical Information and Computer Sciences **41**, 830 (2001), pMID: 11410065, <http://dx.doi.org/10.1021/ci000459c>.
 - [7] T. Chen, J. Morris, and E. Martin, Chemometrics and Intelligent Laboratory Systems **87**, 59 (2007), selected papers presented at the Conferentia Chemometrica 2005 Hajdszoboszl, Hungary 28-31 August 2005Conferentia Chemometrica 2005.
 - [8] M. J. Way, L. V. Foster, P. R. Gazis, and A. N. Srivastava, The Astrophysical Journal **706**, 623 (2009).
 - [9] D. G. Bonfield, Y. Sun, N. Davey, M. J. Jarvis, F. B. Abdalla, M. Banerji, and R. G. Adams, Monthly Notices of the Royal Astronomical Society **405**, 987 (2010), <http://mnras.oxfordjournals.org/content/405/2/987.full.pdf+html>.
 - [10] G. T. von Nessi and M. J. Hole, Review of Scientific Instruments **84**, 063505 (2013), <http://dx.doi.org/10.1063/1.4811378>.

- [11] M. Chilenski, M. Greenwald, Y. Marzouk, N. Howard, A. White, J. Rice, and J. Walk, *Nuclear Fusion* **55**, 023012 (2015).
- [12] I. Yamada, K. Narihara, H. Hayashi, and L. E. Group, *Review of Scientific Instruments* **74**, 1675 (2003).
- [13] I. Yamada, K. Narihara, H. Funaba, R. Yasuhara, T. Kohmoto, H. Hayashi, T. Hatae, H. Tojo, T. Sakuma, H. Yoshida, H. Fujita, and M. Nakatsuka, *Journal of Instrumentation* **7**, C05007 (2012).
- [14] C. Bishop, *Pattern Recognition and Machine Learning* (Springer-Verlag New York, 2006).
- [15] N. J. PEACOCK, D. C. ROBINSON, M. J. FORREST, P. D. WILCOCK, and V. V. SANNIKOV, *Nature* **224**, 488 (1969).
- [16] T. N. Carlstrom, G. L. Campbell, J. C. DeBoo, R. Evanko, J. Evans, C. M. Greenfield, J. Haskovec, C. L. Hsieh, E. McKee, R. T. Snider, R. Stockdale, P. K. Trost, and M. P. Thomas, *Review of Scientific Instruments* **63**, 4901 (1992).
- [17] H. Murmann, S. Gtsch, H. Rhr, H. Salzmann, and K. H. Steuer, *Review of Scientific Instruments* **63**, 4941 (1992).
- [18] B. P. LeBlanc, R. E. Bell, D. W. Johnson, D. E. Hoffman, D. C. Long, and R. W. Palladino, *Review of Scientific Instruments* **74**, 1659 (2003).
- [19] I. Yamada, L. Narihara, H. Hayashi, H. Funaba, and L. experimental group, *Plasma and Fusion Research* **1**, S1106 (2007).
- [20] C. Suzuki, K. Ida, Y. Suzuki, M. Yoshida, M. Emoto, and M. Yokoyama, *Plasma Physics and Controlled Fusion* **55**, 014016 (2013).
- [21] V. Tresp, *Neural Computation* **12**, 2719 (2000).
- [22] C. Holland, A. E. White, G. R. McKee, M. W. Shafer, J. Candy, R. E. Waltz, L. Schmitz, and G. R. Tynan, *Physics of Plasmas* **16**, 052301 (2009), <http://dx.doi.org/10.1063/1.3085792>.
- [23] C. K. I. W. Edwin V. Bonilla, Kian Ming A. Chai, *Proceedings of NIPS2007* (2007).